Supplementary Note 1. Extending *diseaseQUEST* to other model organisms and diseases.

In this study, we have used the *C. elegans* model system as a proof of principle for the construction of *in silico* functional network representations of model organism biology. Here, we describe how one can apply *diseaseQUEST* to other model organisms (e.g., mouse, fly, zebrafish) and diseases of interest.

A. Functional Representation module

Below, we describe how to apply the Functional Representation module to other organisms (e.g., mouse, fly, zebrafish).

Data integration process

We have provided a *diseaseQUEST* docker image (https://github.com/FunctionLab/diseasequest-docker) that uses the Sleipnir functional library to construct the *in silico* functional network models for the model organism of interest. Currently, we provide a small worm test data compendium for users to try that is automatically downloaded upon setup in the `data/` directory.

The files relevant to network integration are located in `./data/network_integration/`. To construct networks for another model organism of interest, the user simply has to replace the `genes.txt` file with a list of genes from the corresponding organism, the files in `./data/network_integration/data_compendium/` with the data compendium assembled in the "Data compendium assembly" section below, and the files in `./data/network_integration/gold_standard/` and `./data/network_integration/weights/` with the files generated from the "Gold standard construction" section below (parts iii and iv).

Then, by running the following command:

```
docker run -v `pwd`/data/:/dq/data -v `pwd`/outputs:/dq/outputs dq networks [tissue name]
```

the corresponding tissue network for the model organism of interest will be generated and located in `./outputs/all/predictions`.

Data compendium assembly

Below we provide guidance on assembling a data compendium for any model organism of interest, to be used with the user-friendly *diseaseQUEST* docker image. We have shown (Supplemental file 8) that the network construction method is robust to data compendia size, but predictive performance does still improve with more data, so it is important to assemble a large genome-wide data compendium.

(i) Gene expression data

Gene expression datasets for a large number of model organisms of interest are available for download from the Gene Expression Omnibus (GEO) data repository maintained by NCBI (https://www.ncbi.nlm.nih.gov/geo/).

(ii) Physical interaction data
For most model organisms of interest, there have been systematic physical interaction studies, and these data can be downloaded from relevant databases, with the main two sources being BioGRID (https://thebiogrid.org) and IntAct (https://www.ebi.ac.uk/intact/).

(iii) Transcription factor binding profiles
Shared transcription factor binding profiles are also informative for analysis of functional similarity between genes. Experimentally defined transcription factor binding sites can be downloaded from the JASPAR database (http://jaspar.genereg.net), and the 1kb upstream regions of each gene in the genome for the model organism of interest can be scanned for the presence of transcription factor binding site motifs using the MEME software suite (http://meme-suite.org). The Fisher z-transformed Pearson correlation of these profiles is a good measure of similarity.

(iv) Organism-specific data (e.g., genetic interaction data)
There are still other types of data from different types of screens and experimental assays that may be available in different model organisms that may capture the functional similarity of genes. One such example is genetic interaction screens in worm and fly. The Fisher z-transformed Pearson correlation of these profiles is a good measure of similarity.

Gold standard construction
(i) Global functional standard
Biological process annotations for the model organism of interest can be downloaded from the Gene Ontology (http://www.geneontology.org). After propagating all annotations with experimental evidence codes (i.e., EXP, IDA, IPI, IMP, IGI, IEP), genes co-annotated to the same slim term are considered positive examples in the gold standard. Genes lacking co-annotations to any term or only co-annotated to highly overlapping or high-level GO terms are considered negative examples.

(ii) Tissue-gene expression standard
For most model organisms of interest, each community has curated known tissue gene relationships based on small-scale expression analyses and should be downloaded from the corresponding resource. (Exclusion of results from microarray or RNA-seq results here are to ensure specificity and quality). For example, the Gene eXpression Database (http://www.informatics.jax.org/expression.shtml) maintained by MGI provides mouse tissue expression annotations, FlyAtlas (http://flyatlas.org) is a database of fly tissue expression, WormBase (https://wormbase.org) curates worm tissue-gene expression, and ZFIN (https://zfin.org) has curated zebrafish tissue-gene expression. Typically, we recommend only constructing networks for tissues with at least 10 direct gene annotations (meaning they are sufficiently well understood).

(iii) Incorporating tissue-specificity into functional gold standard
The description in the corresponding Methods section is model organism agnostic and can be directly applied.

(iv) Supplementation of tissue-specific gold standard using previously unlabeled features
The description in the corresponding Methods section is model organism agnostic and can be directly applied.

B. Disease Prediction module

In our provided docker image (see description above), we have also provided functionality that leverages the Sleipnir library to make disease gene predictions. After assembling the gold standard ("Assembling human disease gold standards" section below), the file with the gold standard (tab-delimited file with gene name and 1 for positive example, -1 for negative example, see example file in docker image) can be placed in `./data/disease_prediction`.

Then, running the following command (with the relevant tissue functional representation for the model organism of interest generated above):

```
docker run -v `pwd`/data/:/dq/data -v `pwd`/outputs:/dq/outputs dq predictions
[disease name] [tissue name]
```

will generate the corresponding set of disease predictions for the model organism in `./outputs/`.

Assembling human disease gold standards

To make disease predictions for the disease of interest, a gold standard of genes identified from related quantitative genetics studies needs to be compiled. For example, in our study, we downloaded genes reported in various GWAS studies from the GWAS Catalog (https://www.ebi.ac.uk/gwas/). Note that while we use the GWAS Catalog here as a large database of quantitative genetics studies, other compilations (e.g., Simons Simplex Collection for autism (https://www.sfari.org/resource/simons-simplex-collection/)) can be used as well. After identifying the functional analogs of these reported genes in the model organism of interest as positive examples, orthologs of other genes implicated by quantitative genetics studies (for other diseases) can be used as negative examples.

C. Phenotypic Assay module

The choice of experimental screen for further prioritization of the disease predictions will be dependent on the disease and model organism of interest and chosen by the expert biologist applying *diseaseQUEST*. For example, a study of Alzheimer's Disease using mouse entorhinal cortex functional representations could be further screened using a novel object recognition assay. Alternatively, a heart rate assay in zebrafish could be used to further screen for hypertension gene candidates that were predicted from the cardiac arrhtymia GWAS genes using zebrafish functional representations of the pronephron.